Application note

Simplified targeted genotyping by sequencing using the Amp-Seq Reagent System and the Complete Genomics DNBSEQ-T7 sequencer

Introduction

Targeted genotyping by sequencing (targeted GBS) is a powerful tool in agribiotechnology, driving advancements in food security and animal genetics by enabling precise identification of genetic variants in plants and animals. By focusing on specific regions of the genome, this application allows the detection of key markers linked to traits of interest, such as disease resistance, drought tolerance, and increased yield. Compared to whole genome sequencing (WGS) methods, targeted GBS is growing in adoption at many agricultural biotechnology companies due to its cost-effectiveness and scalability, enabling the simultaneous analysis of large numbers of single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). An increase in demand for targeted sequencing methods is anticipated as new next generation sequencing (NGS) companies drive down sequencing costs. LGC Biosearch Technologies™ have developed a method to simplify and streamline targeted genotyping workflows, ensuring ease of adoption and scalability to industrial throughput. This application note highlights the combination of LGC's Amp-Seq, a novel technology for targeted GBS sample preparation, with sequencing using the Complete Genomics DNBSEQ-T7 sequencing platform. It explores how this new technological combination reduces cost per sample, is easily automatable for high-throughput genotyping of crop samples and produces high-quality, uniform data.

BIOSEARCH™ TECHNOLOGIES

Application note

Simplified targeted GBS using the Amp-Seq Reagent System and the Complete Genomics DNBSEQ-T7 sequencer

Methods

Library preparation

LGC's Amp-Seq design pipeline was used to develop an Amp-Seq panel targeting more than 1,000 genomic loci in soy (SNPs and indels) as specified by a customer in the Amp-Seq pilot program. DNA was purified from customer-provided soy leaf samples (n=384) using LGC's sbeadex technology. The purified samples ranged in concentration from 2.5 to 50 ng/µL, with an average of 20 ng/µL. Samples were diluted 20-fold to an average concentration of 1 ng/µL, maintaining a sample concentration range of 0.125 to 2.5 ng/µL.

Amp-Seq libraries were produced via the 2-stage method (see figure 1) as described in the Amp-Seq Reagent System Manual; the manual can be requested using the form at the bottom of this webpage. Stage 1 reactions (5 µL) were assembled by dispensing 4 µL of a pre-mix containing Amp-Seq High Sensitivity Amplification Master Mix and the design-specific Amp-Seq Stage 1 reagent into each well of a 384-well plate. The diluted DNA samples (1 µL) were added, and the Stage 1 reactions were amplified in a thermocycler for 24 cycles using an annealing/extension temperature of 66 °C.

During the Stage 1 cycling program, a Stage 2 reaction pre-mix containing the Amp-Seq High Sensitivity Amplification Master Mix and the design-specific Amp-Seq Stage 2 reagent was assembled and dispensed into each well of the provided single-use 384-well Amp-Seq Dried Index plate (4 µL per well). When Stage 1 amplification was complete, 1 µL of each Stage 1 product was transferred to the Stage 2 reaction plate, and the Stage 2 reactions were amplified using the same cycling conditions as for Stage 1.

Once the Stage 2 amplification was completed, all 384 reactions were pooled via centrifugation into a collection trough. A 300 µL aliquot of the pool was subjected to two rounds of cleaning with AMPure XP paramagnetic beads (Beckman), reusing the beads by adding the PEG/NaCl solution provided with the Amp-Seq kit. The final cleaned library pool was eluted into 22 µL of the provided Low TE buffer and quantified using a Qubit dsDNA HS kit (Invitrogen). Library products were evaluated for size and the presence of primer-dimer products on a High-Sensitivity Bioanalyzer 2100 chip (Agilent). The final yield of clean library was 47.5 ng/µL, with an average size of 425 bp. Small primer-dimer sized products (~150-170 bp) comprised <1% of the total product.

The bead-cleaned and quantitated library pool was sent to Complete Genomics for library conversion and sequencing on the DNBSEQ-T7 sequencing platform.

Application note

Simplified targeted GBS using the Amp-Seq Reagent System and the Complete Genomics DNBSEQ-T7 sequencer

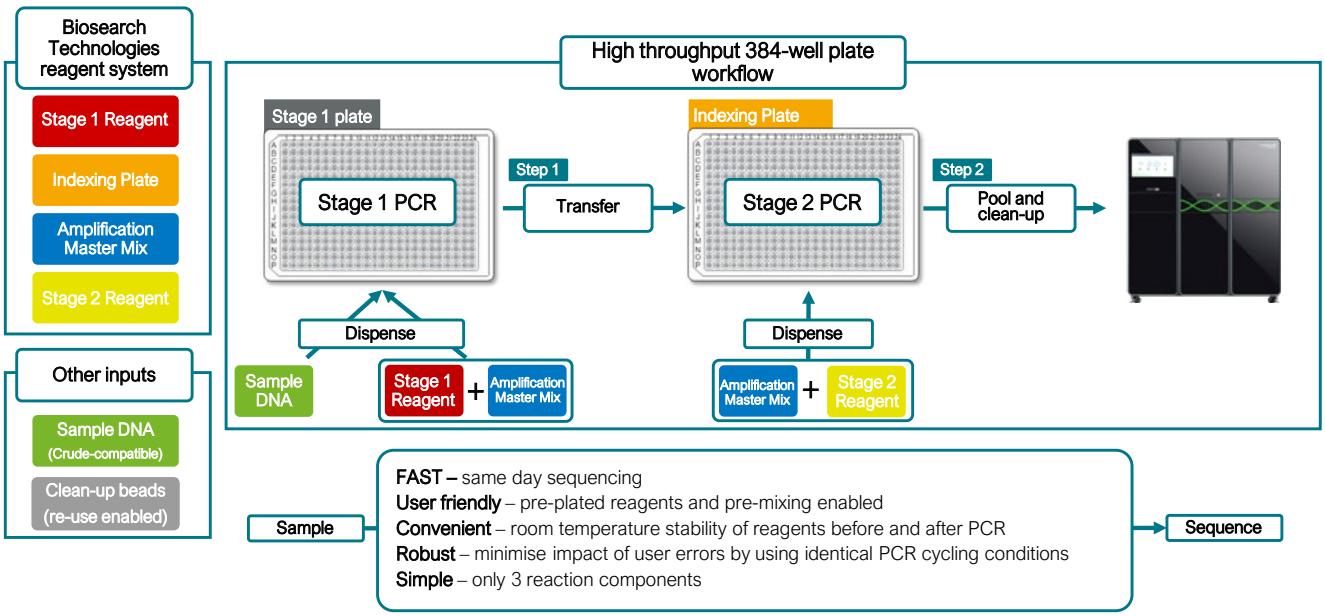

Figure 1. Schematic of the Amp-Seq Reagent System automated workflow incorporating the DNBSEQ-T7 sequencing platform.

Library conversion

The first step in the library preparation process utilises existing i5 and i7 adapters to convert the samples into a single stranded library while incorporating the appropriate CG adapters. Typically, the libraries are digested, and a single strand of template is circularized and amplified to create a DNBSEQ library using the CG Universal Library Conversion Kit (App-A, Cat. No 1000004155). This process was performed using the Complete Genomics DNBSEQ Universal Library Conversion Kit. Following conversion, the libraries were pooled, circularized and made into DNA Nanoballs (DNBs) for sequencing on the T7. The run chemistry was configured as a custom 2x150 bp run, with reduced data output generated for the second read index.

Bioinformatics analysis

The sequencing data was demultiplexed using the i5 and i7 indices into individual filtered FASTQ files. Data analysis was performed using the custom data analysis software BiosearchCaller, developed by LGC specifically for analysing targeted sequencing data from the Amp-Seq Reagent System (the manual can be requested using the form at the bottom of this webpage). This software calls genotypes for SNP and short indel (up to 10 bp) targets from diploid and polyploid species.

Additionally, BiosearchCaller supports de novo discovery of variants from targeted amplicons. To enhance the data analysis quality, input data, including marker details, such as chromosome number and coordinate positions, sequencing read files, and the reference genome, can be provided to BiosearchCaller to improve genotype calls as requested by the user.

Application note

Simplified targeted GBS using the Amp-Seq Reagent System and the Complete Genomics DNBSEQ-T7 sequencer

Results

The sequencing results obtained show that the Complete Genomics DNBSEQ-T7 sequencing platform generated more than 6 billion high quality reads in 25 hours with 98% of these reads retained after demultiplexing (table 1). With over 97% of reads exhibiting a quality value >Q30, the data quality exceeds the highest requirements for Amp-Seq product specifications (table 2).

Category	Value
Software version	BSC_1.4.2.177
Template version	0.8.0
Reference	NULL
Cycle number	171
Chip productivity (%)	71.94
Image area	1764
Total reads (M)	6277.62
Q30 (%)	97.17
Split rate (%)	98.00
Runon1 (%)	0.06
Runon2 (%)	0.05
Lag1 (%)	0.12
Lag2 (%)	0.63
ESR(%)	75.45

Table 1. Sequencing output from the DNBSEQ-T7 run with the Amp-Seq amplicons. The number of total reads are reported in the millions (M) and the SplitRate (%) metric shows the percentage of reads successfully demultiplexed.

Category	Number of reads	Number of bases	Q10%	Q20%	Q30%	EstErr%	GC%	N%
Read 1	6,277,618,276	941,642,741,400	99.73	99.12	97.25	0.15	47.61	0
Read 2	6,277,618,276	31,388,091,380	99.33	98.07	94.78	0.32	50.69	0
Total reads	6,277,618,276	973,030,832,780	99.72	99.08	97.17	0.15	47.7	0

Table 2. Summary of the fastq statistics from the individual reads. Definitions of statistics as follows: number of reads (total reads in read 1 and read 2 fastqs), number of bases (total bases in read 1 and read 2 fastqs), percentage of guanidine and cytosine bases (GC%), quality values (Q10% = 90%, Q20% = 99%, Q30% = 99.9%) and the estimated error rate (EstErr%).

Application note

Simplified targeted GBS using the Amp-Seq Reagent System and the Complete Genomics DNBSEQ-T7 sequencer

Figure 2 shows the performance metrics of the Amp-Seq products when sequenced on the DNBSEQ-T7 platform as compared to the performance of these same Amp-Seq products when sequenced on a leading competitor platform. Genotype data concordance between the two platforms was >99% demonstrating the high quality of reliable data generated by the DNBSEQ-T7. Both platforms performed nearly identically across all sequencing data metrics for this Amp-Seq data, with the DNBSEQ-T7 results displaying higher percentage mapping to genome and on-target mapping rates.

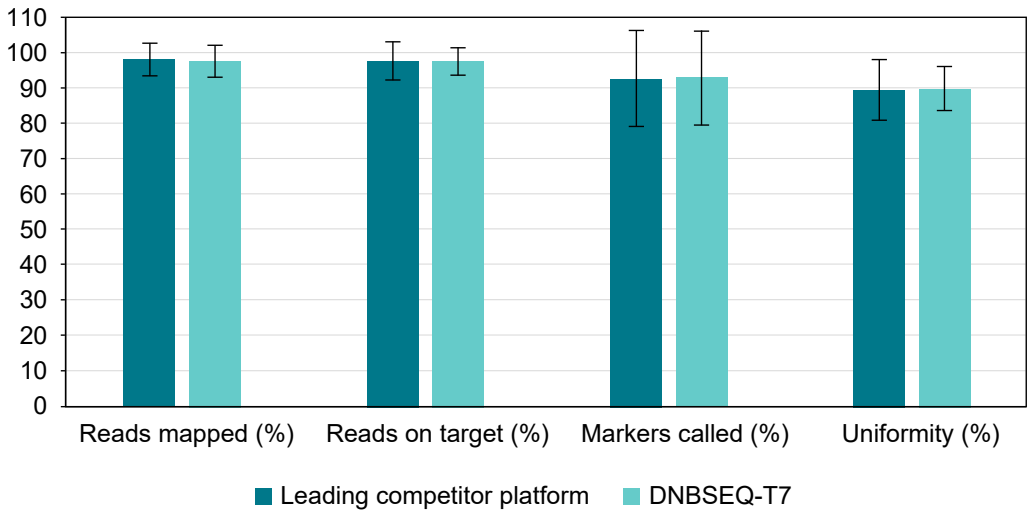

Figure 2. Dataset for an Amp-Seq 1000+ marker panel for soy. Across all metrics, comparable performance is demonstrated between the DNBSEQ-T7 and the leading competitor platform.

Summary

Here we have presented the successful combination of LGC's Amp-Seq technology for targeted GBS with the Complete Genomics DNBSEQ-T7 sequencer, achieving high genotyping call rates and industry-leading sequencing data quality. This pairing of technology and platform offers a valuable solution to the agribiotech sector, with the scalability and affordability of Amp-Seq and the time and cost savings for generation of high-quality sequencing data.

BIOSEARCH™ TECHNOLOGIES